DOCUMENT RESUME

ED 408 308 TM 026 521

AUTHOR Vidal, Sherry
TITLE Canonical Correlation Analysis as the General Linear Model.
PUB DATE Jan 97
NOTE 35p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, January 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Correlation; Heuristics; *Multivariate Analysis; *Regression (Statistics); Satisfaction; Synthesis
IDENTIFIERS *F Test; *General Linear Model

ABSTRACT

The concept of the general linear model (GLM) is illustrated and how canonical correlation analysis is the GLM is explained, using a heuristic data set to demonstrate how canonical correlation analysis subsumes various multivariate and univariate methods. The paper shows how each of these analyses produces a synthetic variable, like the Yhat variable in regression. Ultimately these synthetic variables are actually analyzed in all statistics, a fact that is important to researchers who want to understand the substance of their statistical analysis. The illustrative (fictitious) example involves the relationship between a set of marital happiness characteristics, including a marital satisfaction score and a frequency of sex score, and a set of personal characteristics, which includes IQ scores and overall religiosity. The latent constructs, marital happiness and personal characteristics, are the sets of variables that are examined. A brief summary of the canonical correlation analysis is presented, and how canonical correlation subsumes regression, factorial analysis of variance, and T-tests is discussed. The discussion makes it clear that the "F" statistic is not the sole statistic of interest to researchers. The use of canonical correlation as GLM can help students and researchers comprehend the similarities between models as well as the different statistics that are important in all analyses, such as synthetic variables. (Contains 6 figures, 7 tables, and 18 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 408 308

Running head: CANONICAL CORRELATION ANALYSIS AS THE GENERAL LINEAR

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
SHERRY VIDAL
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Canonical Correlation Analysis as the General Linear Model

Sherry Vidal

Texas A&M University 77843-4425

Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX, January 1997.

TM026521

BEST COPY AVAILABLE

Abstract

The present paper illustrates the concept of the general linear model (GLM) and how canonical correlational analysis is the general linear model. Through a heuristic data set how canonical analysis subsumes various multivariate and univariate methods is demonstrated. Furthermore, the paper illustrates how each of these analyses produce a synthetic variable, like the Yhat variable in regression. Ultimately it is these synthetic variables are actually analyzed in all statistics and which tend to be of extreme importance to erudite researchers who want to understand the substance of their statistical analysis.

Canonical Correlation Analysis as a General Linear Model

Many graduate students, like the author, often learn statistics with a relatively limited conceptual understanding of the foundations of univariate and multivariate analyses. Maxwell, Camp, and Arvey (1981) emphasized that "researchers are not well acquainted with the differences among the various measures (of association) or the assumptions that underlie their use" (p. 525). Frequently, many researchers and graduate students make assertions such as "I would rather use Analysis of Variance (ANOVA) than regression in my study because ANOVA is simpler and it will provide me with all the information I need." Comments such as these are ill-informed and often result in the use of less desirable data analytic tools. Specifically, all analyses are correlational and produce similar latent variables, however the decision to choose a statistical analysis should not be based on its simplicity, but rather on how the analysis fits with the reality of the data and research model.

Ultimately, all analyses such as the t-test, Pearson correlation, ANOVA, and MANOVA are subsumed by correlational analysis, and more specifically canonical correlation analysis. In 1968 Cohen acknowledged that ANOVA was a special case of regression; he stated that within regression analyses "lie possibilities for more relevant and therefore more powerful exploitation of research data" (p. 426). Cohen (1968) was emphasizing that two statistical analyses could yield the same results, but that one might provide more useful information. Consequently, it is important to have an understanding of the model which subsumes all analyses, this model is called the general linear model, or GLM.

The general linear model "is a linear equation which expresses a dependent (criterion) variable as a function of a weighted sum of independent (predictor) variables" (Falzer, 1974, p. 128). Simply stated, the GLM can produce an equation which maximizes the relationship of the

independent variables to dependent variables. In regression analysis, this equation is called a regression equation. In factor analysis these are called factors, and in discriminant analysis and canonical analysis they are called functions. Figure 1 illustrates how these various synthetic variables, as opposed to observed variables, exist in all statistical analyses.

INSERT FIGURE 1 ABOUT HERE

Moreover, these synthetic variables are the variables that researchers are most interested in evaluating, rather than a specific t or F statistic. The synthetic variables are often evaluated as opposed to the t or F statistic to determine *what the findings are* rather than *if they are statistically significant*. As a result, canonical correlation analysis (CCA) can act as a GLM across these different statistical methods.

The purpose of the present paper is to illustrate the foundations of the general linear model, using canonical correlation analysis, and to provide a heuristic data set to illustrate the correlational link between these analyses. This discussion will be primarily conceptual in nature, and more explicit computational detail can be found in Tatsouka (1975). Although Cohen (1968) and Falzer (1974) acknowledged the importance of the general linear model in the 60's and 70's, the use of ANOVA methods remained popular through the 80's because of their computational simplicity over other methods such as regression. Since computational aids such as high powered computers were unavailable to many researchers until the 1980's, researchers used analytical methods which were congruent with existing technology.

Fortunately, computers today can compute complex analyses such as regression, and canonical analysis, however the shift from OVA methods to the general linear model has been gradual. During the years 1969-1978, Wilson (1980) found that 41% of journal articles in an

educational research journal used OVA methods as compared with 25% during the years 1978-1987 (Elmore & Woehlke, 1988). Researchers are beginning to recognize that the general linear model

> can be used equally well in experimental or non-experimental research. It can handle continuous and categorical variables. It can handle two, three, four or more independent variables.... Finally, as we will abundantly show, multiple regression analysis [and canonical correlation analysis] can do anything that the analysis of variance does—sums of squares, mean squares, F ratios—and more. (Kerlinger & Pedhazur, 1973, p. 3)

Advantages of the General Linear Model

One of the primary advantages of the general linear model is the ability to use both categorical variables and intervally-scaled variables. OVA analyses require that independent variables are categorical, therefore observed variables which are not categorical must be reconfigured into categories. This process often results in a misrepresentation of what the variable actual is. Imagine a fresh batch of chocolate chip cookies where each cookie has a variety of chocolate chips. Often children become excited by the number of chocolate chips that are in each cookie. Next, imagine a world where each batch of chocolate chip cookies resulted in a cookie either containing one chocolate chip or two chips. In such a world, children and adults would no longer be as interested in the variety that chocolate chip cookies provided. Similarly, when a researcher dichotomizes variables, variety (variance) is decreased and this limits our understanding of individual differences. While variation in a cookie is not similar to variations of individuals, this illustration represents how reducing an interval variable (multichip cookie) into a

dichotomy (one chip or two chip cookie) or trichotomy can change the characteristics of a variable (cookie). Pedhazur (1982) stated:

> categorization of attribute variables is all too frequently resorted to in the social sciences... It is possible that some of the conflicting evidence in the research literature of a given area may be attributed to the practice of categorization of continuous variables... Categorization leads to a loss of information, and consequently a less sensitive analysis. (pp. 452-453)

Furthermore, Thompson (1986) has established that ANOVA methods tend to overestimate smaller effect sizes: "OVA methods tend to reduce power against Type II errors by reducing reliability levels of variables that were originally higher than nominally scaled. Statistically significant effects are theoretically possible only when variables are reliably measured" (p. 919). Therefore, the use of a general linear model increases the likelihood that the analysis will be replicable, especially when an interval variable is converted into a categorical variable. Moreover, "multivarite methods such as canonical correlation analysis best honor the nature of the reality that most of us want to study, because most of us believe we live in a reality where most effects have multiple causes and most causes have multiple effects" (Thompson, in press, p. 2).

In conclusion, Arnold (1996) succinctly summarizes the general linear model framework into four main areas as follows:

1) all analyses are correlational and yield a measure of effect size that is analogous to r^2;
2) all parametric techniques invoke least-squared weights [beta weights in regression, canonical function coefficients in canonical correlation analysis, etc.];
3) the general linear model can do anything that the specific models can do; and

4) canonical correlation is the general linear model. (p. 3)

Thus, the general linear model is the conceptual umbrella to understand the links data analytic models. Furthermore, to understand multivariate and univariate analyses it is imperative to comprehend the model which subsumes these analyses.

Overview of Canonical Correlation Analysis

Canonical correlational analysis is very similar to regression, such that there are a set of predictor variables and a set of criterion variables and the researcher wishes to evaluate the relationship between the two sets. However, in canonical analysis each "set" of variables (the criterion set, and the predictor set) represent a latent construct which the researcher is examining. Hotelling (1936) developed canonical correlation analysis to evaluate this type of linear correlation between variables sets. While canonical analysis can consider more than two sets of variables at a time, "most researchers us canonical correlation analysis in situations involving only two variable sets" (Thompson, in press, p. 1). In the present analysis we will be examining the relationship between a set of *marital happiness characteristics*, which includes a marital satisfaction score and a frequency of sex score reported by the couple, compared to a set of *personal characteristics*, which includes an IQ score for females, and IQ score for men, and an overall score of religiosity reported by the couple. The data are reported in Table 1. Please note that the data are fictitious. These two latent constructs, *marital happiness* and *personal characteristics* will be the two "sets" of variables which will be examined.

INSERT TABLE 1 ABOUT HERE

However, notice that it is the five observed variables which are entered as data for the analysis, not two scores on each latent construct set. Canonical correlation analysis is conducted only when variable sets are thought to "exist within meaningful variable sets" (Thompson, in press, p. 6). If the variables do not exist within meaningful sets then canonical analysis would not be an appropriate data analytic tool. In the present example, all of the variables are intervally scaled and appear to create two somewhat meaningful variable sets. Ideally, there should be a ratio of 20 subjects to each variable, however, since this example is for illustration only, this assumption will not be met. It is also important to consider selecting a small number of variables to make the model more parsimonious. One can reduce the number of variables by doing a principal components analysis to compute factor scores which would help the researcher utilize variables which are more representative of the construct one wishes to measure.

The next series of steps in a canonical correlation analysis can get quite complicated, and since the purpose of this paper is not to explore all the mathematics involved, only a brief summary of the computations will be explored. For a more in-depth presentation of the computations please refer to Stevens (1996). For the present paper, SPSS FOR WINDOWS was used to compute the following analyses. The computer program is reported in Appendix A. First, the computer program creates a correlation matrix and then partitions the matrix into quadrants that are related to the variable sets. Thompson (1984, in press) states that a quadruple product matrix is created using the correlation quadrants in the algorithm:

$$R22_{2x2}^{-1} R21_{2x3} R11_{3x3}^{-1} R12_{3x2} = A_{2x2}.$$

Furthermore, Thompson (in press) emphasizes that:

> it is this matrix, A_{2x2}, which is actually then subjected to a principal components analysis, and the principal components results are expressed as standardized weights (Thompson, 1984,

pp. 11-14 provides more detail) called 'standardized canonical function coefficients'. These function coefficients are directly akin to beta weights in regression or the pattern coefficients from exploratory factor analysis. (pp. 6-7)

For the present example, the canonical correlation results are reported in Figure 2. Note the canonical correlation coefficient is equal to .741 (Rc = .741) on the first function and .559 on (Rc= .559)on the second function. However, the coefficient on the last function (Rc=.559) represents a test of a single effect with a given function. The first test statistic (Rc=.741) is a test of the *set* of all possible effects (Thompson, in press). The squared canonical correlation coefficient (Rc^2) is an effect size measure. Standardized function coefficients are also reported for the two functions.

INSERT FIGURE 2 ABOUT HERE

Using the function coefficients one can easily apply these weights similarly in a regression analysis. Thompson (in press) presents this in further detail. Similar to regression, in canonical analysis it is also important to evaluate the structure coefficients for each of the variables. Although these weights are not reported in this printout, structure coefficients can be computed by taking the product moment correlation of the measured variable with the synthetic variables. Recall, that suppression effects can occur when a low function coefficient is reported, but the structure coefficient is fairly high. Thompson (1984) presents a thorough discussion on the importance of evaluating both structure and function coefficients. Most importantly, if one fails to examine both of these coefficients, then erroneous conclusions may be derived.

In addition to canonical correlation coefficients, function coefficients, and structure coefficients, there are many other statistics which can be evaluated in a canonical analysis. However, the present paper will focus on canonical coefficients (Rc, Rc^2), function coefficients, and the Wilks lambda (λ) which is also reported on the SPSS printout. Lambda represents an effect size measure and this value is equal to $1-r^2$. The relationship between these statistics and other analyses, such as regression, will be presented later.

Canonical Correlation Analysis as the General Linear Model

Nonetheless, you might still be wondering how does canonical correlation analysis act as a general linear model. Knapp (1978) stated that "virtually all of the commonly encountered tests of significance can be treated as special cases of canonical correlation analysis" (p. 410). Therefore, let's examine how canonical analysis subsumes regression, factorial ANOVA, and T-tests. Through this illustration it is hoped that the reader will realize that all analyses are correlational and that canonical analysis is the general linear model.

Regression and CCA

To illustrate that canonical analysis subsumes regression, only one dependent (criterion) variable, the marital satisfaction score, and three "predictor" variables (iq-male, iq-female, and religiosity) will be used. See Figure 3 for the abridged SPSS print out which illustrates the regression output using these four variables. An R squared value of .318 (R^2=.318) is reported as well as the standardized coefficients called Betas. The Beta coefficient for IQF is -.181 (β=-.181), for IQM is .360(β=.360), and for Relig is .457(β=.457). Next, refer to Figure 3, this is the abridged canonical printout from SPSS.

The Wilks Lambda of .682 (λ = .682) is reported, and the canonical function coefficients for each criterion variable are also reported. Listed in Table 2 is a comparison of these two results.

INSERT FIGURE 3 AND TABLE 2 ABOUT HERE

Note that the R square value in regression is identical to the Rc^2 value using a canonical analysis. Thus, in this respect the analyses produce identical results. In canonical correlation analysis the standardized coefficients (i.e. weights) are called *standardized function coefficients* rather than beta weights, even though these coefficients are the same not withstanding their different names. Since function coefficients are a bit different mathematically from Beta weights (i.e. are arbitrarily scaled differently), a conversion must be performed to illustrate the relationship between function coefficients and Beta weights. In Table 4 these simple conversions are illustrated. The mathematical conversion is Beta/Canonical correlation coefficient (Rc) or multiply the function coefficient by R. In Table 3, the beta weight for each predictor variable is divided by the canonical function coefficient for that analysis.

INSERT TABLE 3 ABOUT HERE

Thus, it is empirically evident that regression and canonical analyses produce identical results with regards to effect sizes and that the weights share a relationship. While in regression we get one set of weights since there is only one dependent variable, in canonical there is a set of weights for each function. Ultimately, the total number of functions is dependent upon the lowest number of variables in a "set". Therefore, if there are two variables in one set and four in the other, there will be a maximum of two functions and two sets of weights for each function. Of course, when one set consists of only one variable, both regression and CCA yield only one equation (i.e., set of weights).

ANOVA and CCA

Illustrating the relationship between canonical correlation analysis and factorial ANOVA is a bit more complex. To show the relationship between all the pieces in ANOVA, the main effects and the interaction effects, orthogonal contrast variables must be created. Furthermore, the continuous variables must be dichotomized, reconfigured from an interval scale to a categorical. This procedure is not recommended in research, but is conducted here solely for the heuristiic illustration of CCA as the GLM. The variables to be included in this analysis will be the marital satisfaction score (MSS), ovaiqf (a dichotomized IQ score for females) and religova (religiosity dichotomized). The ANOVA analysis from SPSS is printed in Figure 5 and summarized in Table 4. Next, orthogonal contrast variables must be created to conduct the canonical analysis. These new variables will be named "A1", "B1", and "A1B1". "A1" represents ovaiqf where a negative one represents the "lower IQ group" and positive one represents the "higher IQ group". Recall, when creating orthogonal contrast variables, that the sum must equal zero. In the present case this is true. The variable "B1" is also an uncorrelated contrast variable, but it represents the religiosity variable, "and A1B1" represents the 2-way interaction between these two variables, IQF and religiosity.

As stated previously, the relationship between factorial ANOVA and CCA is less obvious. In the canonical analysis a series of analyses must be conducted to obtain the corresponding main effects and interaction effects that ANOVA creates. This is done through a process of using four models, an omnibus test, a test without A1, a test without B1, and a test with no interaction effect. These four models are reported in Table 5 with the corresponding Wilks lambda's reported. Recall, that the Wilks lambda is similar to an effect size.

INSERT TABLE 5 ABOUT HERE

Rao (1952) illustrates how the Wilks lambda also shares a relationship with the F statistic in ANOVA through the following formula:

[1-lambda/lambda] * (df error/df effect) = F statistic.

However, before this formula can be applied to the lambdas a specific source of variance must be calculated for each main effect and interaction. Recall, that the models do not represent the A1 main effect, or the B1 main effect, or solely the interaction effect, thus to acquire an A1 main effect, the omnibus test (model 1) must be divided by the test with no A1 (model 2). This results in a lambda of .9362 (λ=.9362) for the A1 main effect. It is this lambda statistic that can be applied to the above formula. The remaining lambda conversions are reported in Table 6.

INSERT TABLE 6 ABOUT HERE

Inserting the lambda (λ = .9362) into the above formula the resulting F statistic is shown below:

[1- .9362/.9362] * (16/1) = 1.09.

This is the exact same F statistic reported for the OVAIQF in Figure 4 in the ANOVA printout. Figure 5 illustrates the canonical statistics printed from SPSS. As expected, the F statistics for the other main effect and interaction are also identical and are reported in Table 7.

INSERT TABLE 7 ABOUT HERE

CCA and t-tests

Although it should be evident that CCA subsumes univariate and multivariate analyses, one last presentation of how CCA subsumes t-tests will be presented. Figure 6 reports the results from a t-test and canonical correlation using the variables religova and mss. Since a t-test is restricted to the comparison of two means, these two variables were selected. For a t-test, often it is the t value that is evaluated. In this example, t = -2.484. Tatsuoka (1975) illustrated how the t value is simply a function of the correlation coefficient in the following formula:

$$t = \left| r\sqrt{N-2} / \sqrt{1-r^2} \right|$$

Thus, there must be some type of relationship to canonical correlation analysis, since all analyses are correlational! Refer to the t-test and ANOVA results reported in Figure 6.

INSERT FIGURE 6 ABOUT HERE.

The relationship between ANOVA and t-test is illustrated in the F and t statistic printed in the SPSS printout in Figure 6. Recall that $t^2 = F$ (Tatsuoka, 1975). If the t value of -2.48 is squared, then it equals the F statistic 6.17. Therefore, CCA produces the exact same results as a t-test.

Conclusion

This paper has presented some of the basic concepts regarding canonical correlation analysis and how CCA subsumes other analyses. Furthermore, the present paper has illustrated that the F statistic is not the sole statistic of interest to researchers. The use of canonical

correlation as a general linear model can help students and researchers to comprehend the similarities between the models as well as the different statistics that are of importance in all analyses, such as synthetic variables.

Ultimately, statistical models should aid researchers to understanding their data, rather than constrict or change the reality of the measured variables. Thus, the present paper identified how some analyses may be better that others, such as regression versus ANOVA. Furthermore, the present paper portrayed that all statistical analyses are correlational, even though some research designs may not be. *This implies that r^2 effect sizes are available in all analyses, and should always be reported.* The onus is on the researcher to understand the limitations and similarities between research models, thus it is important that instructional tools, such as the general linear model, be used to aid in this understanding.

References

Arnold, M. (1996, January). The relationship of canonical correlation analysis to other parametric methods. Paper presented at the annual meeting of the Southwestern Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 395 994)

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Elmore, R., & Woehlke, P. (1988). Statistical methods employed in the American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1987. Educational Researcher, 17(9), 19-20.

Falzer, P. (1974). Representative design and the general linear model. Speech Monographs, 41, 127-138.

Hotelling, H. (1935). The most predictable criterion. Journal of Experimental Psychology, 26, 139-142.

Kerlinger, F. N., & Pedhazur, E. J. (1973). Multiple regression in behavioral research. New York: Holt, Rinehart, and Winston.

Knapp, T.R. (1978). Canonical correlation analysis: A general parametric significance-testing system. Psychological Bulletin, 85, (2), 410-416.

Maxwell, S., Camp, C., & Arvey, R. (1981). Measures of strength of association: A comparative examination. Journal of Applied Psychology, 66(5), 525-534.

Pedhazur, E.J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd ed.) New York :Holt, Rinehart, and Winston.

Rao, C. R. (1952). Advanced statistical methods in biometric research. New York: Wiley.

Statistical Package for the Social Sciences (SPSS) [Computer Software]. (1995). Chicago:IL SPSS Inc.

Stevens, J. (1996). Applied multivariate statistics for the social sciences (3rd ed.). Mahwah, NJ: Erlbaum.

Tatsuoka, M. (1975). The general linear model: A "new" trend in analysis of variance. Champaign, IL: Institute for Personality and Ability Testing.

Thompson, B. (1984). Canonical correlation analysis: Uses and interpretation. Newbury Park: Sage.

Thompson, B. (1986). ANOVA versus regression analysis of ATI designs: An empirical investigation. Educational and Psychological Measurement, 46, 917-928.

Thompson, B. (1992, April). Interpreting regression results: Beta weights and structure coefficients are both important. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 344 897).

Thompson, B. (in press). Canonical correlation analysis: Basic concepts and some recommended interpretation practices. In L. Grimm & P. Yarnold (Eds.), Reading and understanding multivariate statistics (Vol. 2). Washington, DC: American Psychological Association.

Willson, V. (1980). Research techniques in AERJ articles: 1969 to 1978. Educational Researcher, 9(6), 5-10.

Table 1

Heuristic Data Set

	MSS	SEX	RELIG	IQM	IQF	OVAIQM	OVAIQF	RELIGOVA
1	50	2	3	93	95	1	1	1
2	20	1	9	85	96	1	1	2
3	30	9	0	99	83	1	1	1
4	80	7	8	95	85	1	1	2
5	75	3	9	98	82	1	1	2
6	60	4	5	95	96	1	1	1
7	39	2	4	85	97	1	1	1
8	45	6	3	87	98	1	1	1
9	34	1	2	82	99	1	1	1
10	69	0	9	80	83	1	1	2
11	72	3	8	130	120	2	2	2
12	85	2	8	117	119	2	2	2
13	49	5	6	118	116	2	2	2
14	35	6	5	106	121	2	2	1
15	25	8	4	118	100	2	2	1
16	87	9	3	112	105	2	2	1
17	91	2	8	103	107	2	2	2
18	53	2	5	104	110	2	2	1
19	49	4	4	100	112	2	2	1
20	67	6	8	113	113	2	2	2

Table 2

Canonical Correlation Subsumes Regression

MSS with IQF, IQM, and RELIG

Canonical Analysis		Regression Analysis	
Rc	.564	R	.564
Squared Rc	.318	R^2	.318
Lambda	.682		
Conversion to F			
([1-.682/.682][16/3])=	2.483	F	2.484
p	.098	*p*	.098

Table 3

Canonical Beta Weights Converted to Function Coefficients

Variable	Beta Weight/Canonical Correlation	= Function Coefficient
IQF	-.181/.564	= .320
IQM	.360/.564	= .638
RELIG	.457/.564	= .810

Table 4

Factorial ANOVA
MSS by categorical variables OVAIQF and RELIG

Source	SOS	df	MS	F	p
Main Effects					
OVAIQF	422.62	1	423	1.09	.312
RELIGOVA	2058.36	1	2058	5.29	.035
2 way interaction effect					
ovaiqf*religova	30.61	1	30	.079	.783
Error	6219.60	16			

Table 5

Canonical Analysis Using 4 Models

Model	Predictors of MSS	Lambda
1. Omnibus	a1 b1 a1b1	.69604
2. No a1	b1 a1b1	.74344
3. No b1	a1 a1b1	.92639
4. No interaction	a1 b1	.69946

Table 6

Recalculation of Lambda for each specific source of variance

Source	Model	Calculation	Lambda
A1	Model 1/Model 2	.696/.743	.936
B1	Model 1/Model 3	.696/.926	.751
A1B1	Model 1/Model 4	.696/.699	.9951

Table 7

Conversion of Canonical Lambda's ANOVA F stats

Source	[1-lambda/lambda] * (df error/df effect) =	F
A1	[1- .936/.936] * (16/1) (.068)(16) =	1.09
B1	[1-.7513/.7513] * (16/1) (.331)(16) =	5.29
A1B1	[1-.9951/.9951] * (16/1) (.004)(16) =	.078

Figure 1

Synthetic Variable Comparison

Regression	**Factor Analysis**	**Canonical Analysis**
Beta weights	Factor pattern coefficients	Stdzd. canonical coefficients
Structure coefficients	Structure coefficients	Structure coefficients
Yhats	Factor scores	Canonical function scores
Equation	Factors	Functions

* Note that although the synthetic variables for t-tests have not been listed, one can create similar synthetic variables.

Figure 2

SPSS Printout of CCA results

* * * * * * A n a l y s i s o f V a r i a n c e -- design 1 * * * * * *

EFFECT .. WITHIN CELLS Regression
Multivariate Tests of Significance (S = 2, M = 0, N = 6 1/2)

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.86164	4.03686	6.00	32.00	.004
Hotellings	1.67186	3.90101	6.00	28.00	.006
Wilks	**.31002**	3.97993	6.00	30.00	.005
Roys	.54890				

Note.. F statistic for WILKS' Lambda is exact.

- -

Eigenvalues and Canonical Correlations

Root No.	Eigenvalue	Pct.	Cum. Pct.	Canon Cor. **(Rc)**	Sq. Cor
1	1.217	72.782	72.782	**.741**	**.549**
2	.455	27.218	100.000	**.559**	**.313**

- -

Dimension Reduction Analysis

Roots	Wilks L.	F	Hypoth. DF	Error DF	Sig. of F
1 TO 2	.31002	3.97993	6.00	30.00	.005
2 TO 2	.68726	3.64035	2.00	16.00	.050

- -

* * * * * * A n a l y s i s o f V a r i a n c e -- design 1 * * * * * *

Standardized canonical function coefficients for DEPENDENT variables

	Function No.	
Variable	1	2
MSS	-.089	.998
SEX	.991	.146

- -

Correlations between DEPENDENT and canonical variables

	Function No.	
Variable	1	2
MSS	-.145	.989
SEX	.996	.089

Figure 2 (cont'd)

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Standardized canonical coefficients for COVARIATES
          CAN. VAR.

COVARIATE           1          2

IQF              -.674      -.458
IQM              1.059       .856
RELIG            -.733       .683

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Correlations between COVARIATES and canonical variables
          CAN. VAR.

Covariate           1          2

IQF              -.020       .178
IQM               .494       .648
RELIG            -.632       .772

* * * * * * Analysis of Variance -- design 1 * * * * * *

Variance in covariates explained by canonical variables

CAN. VAR.  Pct Var DE Cum Pct DE Pct Var CO Cum Pct CO

      1       11.789     11.789     21.478     21.478
      2       10.916     22.705     34.904     56.381

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Regression analysis for WITHIN CELLS error term
Dependent variable .. MSS

COVARIATE            B         Beta    Std. Err.     t-Value   Sig. of t

IQF            -.30836     -.18088         .478        -.645         .528
IQM             .56349      .35955         .443        1.272         .221
RELIG          3.69065      .45690        1.687        2.187         .044

COVARIATE   Lower -95% CL- Upper

IQF             -1.322        .705
IQM              -.375       1.502
RELIG             .114       7.267
Dependent variable .. SEX

COVARIATE            B         Beta    Std. Err.     t-Value   Sig. of t

IQF            -.11251     -.52023         .049       -2.277         .037
IQM             .16384      .82408         .046        3.579         .003
RELIG          -.51957     -.50704         .174       -2.979         .009

COVARIATE   Lower -95% CL- Upper

IQF              -.217       -.008
IQM               .067        .261
RELIG            -.889       -.150
```

Figure 2 (cont'd)

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * * Analysis of Variance -- design 1 * * * * * *

EFFECT .. CONSTANT
Multivariate Tests of Significance (S = 1, M = 0, N = 6 1/2)

Test Name          Value     Exact F Hypoth. DF   Error DF  Sig. of F

Pillais           .01827      .13954       2.00      15.00       .871
Hotellings        .01861      .13954       2.00      15.00       .871
Wilks             .98173      .13954       2.00      15.00       .871
Roys              .01827
Note.. F statistics are exact.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Eigenvalues and Canonical Correlations

Root No.    Eigenvalue        Pct.    Cum. Pct.  Canon Cor.

       1          .019     100.000     100.000        .135

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

EFFECT .. CONSTANT (Cont.)
Univariate F-tests with (1,16) D. F.

Variable   Hypoth. SS   Error SS Hypoth. MS   Error MS          F  Sig. of F

MSS          24.87127 6096.60601   24.87127  381.03788      .06527       .802
SEX            .93725   65.13604     .93725    4.07100      .23023       .638

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

EFFECT .. CONSTANT (Cont.)
Standardized discriminant function coefficients
         Function No.

Variable           1

MSS             .476
SEX             .884

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Correlations between DEPENDENT and canonical variables
         Canonical Variable

Variable           1

MSS             .468
SEX             .879
```

Figure 3

SPSS Printout Regression and CCA Output

Regression

Model Summary[a,b]

Model	Variables Entered	Variables Removed	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	RELIG, IQF, IQM[c,d]		.564	.318	.190	19.52

a. Dependent Variable: MSS

b. Method: Enter

c. Independent Variables: (Constant), RELIG, IQF, IQM

d. All requested variables entered.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2839.144	3	946.381	2.484	.098[b]
	Residual	6096.606	16	381.038		
	Total	8935.750	19			

a. Dependent Variable: MSS

b. Independent Variables: (Constant), RELIG, IQF, IQM

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.760	38.204		.255	.802
	IQF	-.308	.478	-.181	-.645	.528
	IQM	.563	.443	.360	1.272	.221
	RELIG	3.691	1.687	.457	2.187	.044

a. Dependent Variable: MSS

Figure 3 (cont'd)

Canonical Analysis

* * * * * * Analysis of Variance * * * * *

EFFECT .. WITHIN CELLS Regression
Multivariate Tests of Significance (S = 1, M = 1/2, N = 7)

Test Name	Value	Exact F	Hypoth. DF	Error DF	Sig. of F
Pillais	.31773	2.48369	3.00	16.00	.098
Hotellings	.46569	2.48369	3.00	16.00	.098
Wilks	**.68227**	**2.48369**	**3.00**	**16.00**	**.098**
Roys	.31773				

Note.. F statistics are exact.

Standardized canonical **Function coefficients** for DEPENDENT variables

Variable	Function No. 1
IQF	**-.321**
IQM	**.638**
RELIG	**.811**

* * * * * * Analysis of Variance -- design 1 * * * * * *

Correlations between DEPENDENT and canonical variables

Variable	Function No. 1
IQF	.179
IQM	.541
RELIG	.878

Eigenvalues and Canonical Correlations

Root No.	Eigenvalue	Pct.	Cum. Pct.	**Canon Cor. (Rc)**	**Sq. Cor**
1	.466	100.000	100.000	**.564**	**.318**

BEST COPY AVAILABLE

Figure 4

SPSS Printout for ANOVA

ANOVA[a,b]

			Unique Method				
			Sum of Squares	df	Mean Square	F	Sig.
MSS	Main Effects	(Combined)	2700.683	2	1350.341	3.474	.056
		OVAIQF	423.624	1	423.624	1.090	.312
		RELIGOVA	2058.367	1	2058.367	5.295	.035
	2-Way Interactions	OVAIQF * RELIGOVA	30.612	1	30.612	.079	.783
	Model		2716.150	3	905.383	2.329	.113
	Residual		6219.600	16	388.725		
	Total		8935.750	19	470.303		

a. MSS by OVAIQF, RELIGOVA

b. All effects entered simultaneously

Figure 5

Abridged SPSS Printout for CCA results comparing ANOVA

Ominbus Test - CCA

* * * * * * A n a l y s i s o f V a r i a n c e -- design 1 * * * * * *

EFFECT .. WITHIN CELLS Regression
Multivariate Tests of Significance (S = 1, M = 1/2, N = 7)

Test Name	Value	Exact F	Hypoth. DF	Error DF	Sig. of F
Pillais	.30396	2.32911	3.00	16.00	.113
Hotellings	.43671	2.32911	3.00	16.00	.113
Wilks	**.69604**	2.32911	3.00	16.00	.113
Roys	.30396				

Note.. F statistics are exact.

- -

Standardized canonical coefficients for DEPENDENT variables
Function No.

Variable	1
A1	-.399
B1	-.875
A1B1	-.107

* * * * * * A n a l y s i s o f V a r i a n c e -- design 1 * * * * * *

Correlations between DEPENDENT and canonical variables
Function No.

Variable	1
A1	-.476
B1	-.916
A1B1	-.075

- -

Variance in dependent variables explained by canonical variables

CAN. VAR.	Pct Var DE	Cum Pct DE	Pct Var CO	Cum Pct CO
1	35.739	35.739	10.863	10.863

- -

Variance in covariates explained by canonical variables

CAN. VAR.	Pct Var DE	Cum Pct DE	Pct Var CO	Cum Pct CO
1	30.396	30.396	100.000	100.000

- -

BEST COPY AVAILABLE

Figure 5 (cont'd)

CCA for No A1

```
* * * * * * Analysis of Variance -- design 1 * * * * * *

EFFECT .. WITHIN CELLS Regression
Multivariate Tests of Significance (S = 1, M = 0, N = 7 1/2)

Test Name          Value    Exact F Hypoth. DF   Error DF  Sig. of F

Pillais           .25656    2.93328       2.00      17.00       .080
Hotellings        .34509    2.93328       2.00      17.00       .080
Wilks             .74344    2.93328       2.00      17.00       .080
Roys              .25656
Note.. F statistics are exact.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

EFFECT .. WITHIN CELLS Regression (Cont.)
Univariate F-tests with (1,18) D. F.

Variable    Sq. Mul. R  Adj. R-sq.  Hypoth. MS    Error MS           F

B1             .25522      .21385     5.05344      .81925     6.16834
A1B1           .00173      .00000      .03427     1.09810      .03121

Variable     Sig. of F

B1                .023
A1B1              .862

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Standardized canonical coefficients for DEPENDENT variables
          Function No.

Variable            1

B1               .997
A1B1             .072

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * * Analysis of Variance -- design 1 * * * * * *

Correlations between DEPENDENT and canonical variables
          Function No.

Variable            1

B1               .997
A1B1             .082

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

BEST COPY AVAILABLE

Figure 5 (cont'd)

CCA for the no B1 Model.

```
* * * * * * Analysis of Variance -- design 1 * * * * * *

EFFECT .. WITHIN CELLS Regression
Multivariate Tests of Significance (S = 1, M = 0, N = 7 1/2)

Test Name          Value     Exact F Hypoth. DF   Error DF  Sig. of F

Pillais           .07361      .67543       2.00      17.00       .522
Hotellings        .07946      .67543       2.00      17.00       .522
Wilks             .92639      .67543       2.00      17.00       .522
Roys              .07361
Note.. F statistics are exact.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Standardized canonical coefficients for DEPENDENT variables
          Function No.

Variable            1

A1                .993
A1B1              .253

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * * Analysis of Variance -- design 1 * * * * * *

Correlations between DEPENDENT and canonical variables
          Function No.

Variable            1

A1                .968
A1B1              .153

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

BEST COPY AVAILABLE

Figure 5 (cont'd)

CCA for no interaction model

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * * Analysis of Variance -- design 1 * * * * * *

EFFECT .. WITHIN CELLS Regression
Multivariate Tests of Significance (S = 1, M = 0, N = 7 1/2)

Test Name          Value    Exact F Hypoth. DF   Error DF  Sig. of F

Pillais           .30054    3.65221       2.00      17.00       .048
Hotellings        .42967    3.65221       2.00      17.00       .048
Wilks             .69946    3.65221       2.00      17.00       .048
Roys              .30054
Note.. F statistics are exact.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Standardized canonical coefficients for DEPENDENT variables
          Function No.

Variable            1

A1                .390
B1                .882

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * * Analysis of Variance -- design 1 * * * * * *

Correlations between DEPENDENT and canonical variables
          Function No.

Variable            1

A1                .479
B1                .922
```

BEST COPY AVAILABLE

Figure 6

SPSS - T-test Printout & CCA Printout

T-test - Religova & MSS

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
MSS	Equal variances assumed	.166	.689	-2.484	18	.023

CCA - for Religova & MSS

* * * * * * Analysis of Variance -- design 1 * * * * * *

Tests of Significance for MSS using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	6655.13	18	369.73		
REGRESSION	2280.62	1	2280.62	**6.17**	.023
CONSTANT	1277.42	1	1277.42	3.46	.079
(Corrected Model)	2280.62	1	2280.62	6.17	.023
(Corrected Total)	8935.75	19	470.30		

R-Squared = .255
Adjusted R-Squared = .214

BEST COPY AVAILABLE

Appendix A

SPSS commands for printouts

```
title 'cca1.sps'.
set blanks = -9999 undefined warn.
data list
  file = 'a:cca1.dat' fixed records=1 table
  /1 mss 1-2 sex 4 relig 6-7 iqm 8-10 iqf 12-14 ovaiqm 16 ovaiqf 18 religova 20.
missing values mss sex relig iqm iqf ovaiqm ovaiqf (-9999).
list variables=all/cases=500/format=numbered.
execute.
subtitle 'cca subsumes factorial anova'.
anova
  mss by ovaiqf(1, 2) religova (1, 2).
execute.
subtitle '2x2 factorial anova- cca subsumes'.
compute a1=-1.
if (ovaiqf eq 2) a1=1.
compute b1=-1.
if (relig gt 5) b1=1.
compute a1b1=a1*b1.
list variables=mss a1 to a1b1/cases=20/format=numbered.
correlations variables=a1 to a1b1.
subtitle 'b1 CCA subsumes factorial multi-way anova'.
manova a1 b1 a1b1 with mss/
  print=signif(multiv eigen dimenr)/
  discrim(stan cor alpha(.999)) .
subtitle 'b2 cca subsumes factorial multi-way ANOVA '.
manova b1 a1b1 with mss/
  print=signif(multiv eigen dimenr)/
  discrim(stan cor alpha(.999)) .
subtitle 'b3 cca subsumes factorial multi-way ANOVA'.
manova a1 a1b1 with mss/
  print=signif(multiv eigen dimenr)/
  discrim(stan cor alpha(.999)) .
subtitle 'b4 cca subsumes factorial multi-way ANOVA'.
manova a1 b1 with mss/
  print=signif(multiv eigen dimenr)/
  discrim(stan cor alpha(.999)) .
Subtitle 'cca subsumes regression'.
REGRESSION
  /missing listwise
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT mss
  /METHOD=ENTER iqf iqm relig  .
MANOVA
  iqf iqm relig WITH mss
  /PRINT SIGNIF(MULTIV EIGEN DIMENR)
  /discrim(stan cor alpha(.999)) .
subtitle 'cca results from glm in SPSS'.
subtitle 'cca using the manova command in SPSS'.
```

```
MANOVA
  mss sex WITH iqf iqm relig
  /PRINT SIGNIF(MULTIV EIGEN DIMENR)
  /discrim(stan cor alpha(.999)) .
subtitle 'cca subsuming t-test'.
T-TEST
  GROUPS=religova(1 2)
    /VARIABLES=mss
  /CRITERIA=CIN(.95) .
manova
  mss with religova
  /PRINT SIGNIF(MULTIV EIGEN DIMENR)
  /discrim(stan cor alpha(.999)) .
```